MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS – 1963 – A

ADA130800

DOWNLOADING AND POST-PROCESSING OF BIBLIOGRAPHIC INFORMATION WITH THE TIS INTELLIGENT GATEWAY COMPUTER

Isom Harrison, Jr.
Technical Information Department

Viktor E. Hampel
Transportation Systems Research

Richard A. Kawin
Control Data Corporation

Lawrence Livermore National Laboratory

DTIC FILE COPY

DTIC
SELECTE
S
JUL 1 9 1983

A

DOWNLOADING AND POST-PROCESSING OF BIBLIOGRAPHIC INFORMATION WITH THE TIS INTELLIGENT GATEWAY COMPUTER*

Isom Harrison, Jr.
Technical Information Department

Viktor E. Hampel
Transportation Systems Research

Richard A. Kawin
Control Data Corporation

Lawrence Livermore National Laboratory
Livermore, CA 94550

September 1982

ABSTRACT

The ⟨TIS⟩ Intelligent Gateway Computer at the Lawrence Livermore National Laboratory provides authorized users automated access to other information centers, downloading of descriptive information and numerical data, and post-processing of bibliographic citations. Included is the aggregation of extracted information into topical files, the elimination of redundancy, and online review for the creation of annotated relevant sets. Post-processing of the reviewed information can be carried out by permutation of titles, abstracts, and descriptors with statistics (some in graphical form) of their single/multi-term expressions, statistical cross-correlation of data elements, and the creation of concordances and indexes. These tools give new insight into a subject matter or the characteristics of corporate/personal publications. These self-guided procedures can be performed online from remote terminals by telephone dial-up, WATS-lines, over TYMNET, and via the ARPA computer network. The TIS Intelligent Gateway Computer permits the linking of terminals among users. Information specialists and information requestors may jointly view and discuss the progress of an interactive search and its analysis from any location. Uncertain legal constraints by commercial information vendors limit the use of downloading and post-processing at this time to bibliographical information in the public domain, e.g. DOE/RECON.

1. INTRODUCTION

More than 1244 bibliographic and numeric data files are now available from 203 online information vendors.[1] This makes it increasingly difficult to identify relevant citations in a unified manner and to extract meaningful information for decision-making.

Most online bibliographic information is still being sold by offline printing following a search. At best, the citations are shown or printed in chronologically reverse order - last publication first. When a search is carried out in a comprehensive and retrospective manner, the end-user is faced with piles of printouts containing redundant citations in different formats from different vendors. Usually, there are no indexes or contents lists to the returns. Manual review and organization of the material is required. Much of the information thus received is probably being discarded unused.
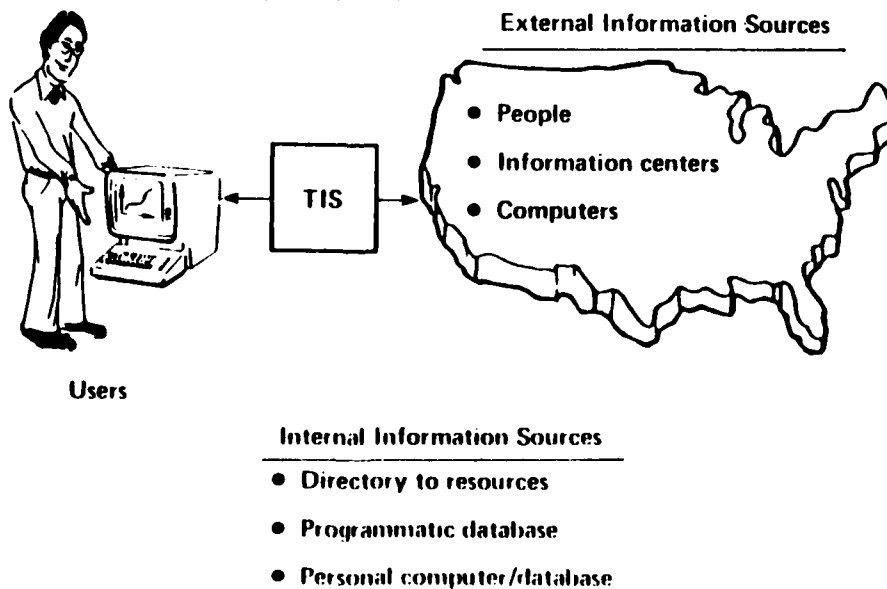
At the Lawrence Livermore National Laboratory (LLNL) we have developed self-guided programs by which some of these tasks can be carried out automatically with a dedicated information machine, the Technology Information System (TIS).[2-4] This system, under development and use since 1975, contains an expanding master directory to databases of other information centers. Authorized users are connected to the named information center automatically and can download desired information to TIS. The resulting files can then be post-processed by programs that permit online review, the display of statistics, the creation of indexes and concordances, and text analysis. In view of the uncertain legal and monetary implications of these powerful procedures, we have limited our applications to the DOE/RECON online information system of the Department of Energy, and are exploring its extensions to other federal information systems.

2. AUTOMATED ACCESS PROCEDURES

We are developing the automated and transparent access procedures to different information centers as part of the prototyping of an Intelligent Gateway Computer (IGC). Users of TIS may consult the availability of programmatic resources stored internally for their use, or made available to them by external information sources.

THE TECHNOLOGY INFORMATION SYSTEM "TIS"

(An Intelligent Gateway Computer)

External Information Sources

- People
- Information centers
- Computers

TIS

Users

Internal Information Sources

- Directory to resources
- Programmatic database
- Personal computer/database

Each external information center is qualified online on TIS by its accredition, the availability and cost of its databases, an annotated description of salient commands and prevailing up-times. This information is extracted by periodic transfer from such centers to TIS for online consultation by the TIS user community prior to their use, which saves time and communication costs. Access to federal and commercial resources is granted to TIS users on an individual basis by the Database Administrator, where appropriate. Authorized users gain access to other information centers simply by giving the command CONNECT, followed by the target name of the desired resource: e.g.,

CONNECT DOE/RECON12   will establish access to DOE/RECON at 1200 baud.

Alternately, users may specify the TIS option number of the desired resource, which is part of each online menu. In either case, users need not be familiar with telephone dial-up numbers, accounts, passwords, or peculiar protocols.

The seven main user communities of the Technology Information System establish their own views of their internal and external programmatic resources, in a self-guided manner, without programmer intervention. Database Administrators assigned to each user community control access rights on a need-to-know/need-to-use basis. Individual users see only those resources (e.g. datafiles, interactive models, graphs and reports) to which they have access. An exception is the external resources we advertise to promote their use. When a user is finished using a TIS-provided external resource, his access rights can be removed by the TIS Database Administrator and no change of passwords is required since none were disclosed.

The DIAL command provides an equally powerful, but user-controlled, method for accessing other information centers and computers. In this case, the user is prompted to specify the telephone number, baud rate, parity, and other parameters, i.e.:

	Telephone	Baud rate	Parity	Duplex	Protocol
		[-300]	[-o]	[-h]	[-b]
DIAL	[number]				
		[-1200]	[-e]	[-f]	[-v]

TIS then establishes the communication similar to an automated telephone dialer. Users have to provide their own accounts and passwords for login on the external host machine. Such procedures can be saved for personal, routine use.

When an account with another information center is opened for TIS, the vendor bills TIS at LLNL, which, in turn, deducts the appropriate costs from the responsible programmatic accounts. When users establish their own accounts with information centers and use them via TIS, they are billed directly by the vendors, who cannot distinguish by what means the user accessed and used their information center.[5]

3. DOWNLOADING OF INFORMATION

The user can initiate downloading of information from another system to TIS in two ways: First, the SAVEON option permits extraction of information when used with the CONNECT command discussed previously. In this case, all the information seen on the screen during one session is placed into a user-named file.

POSTPROCESSING OF BIBLIOGRAPHIC INFORMATION

Conventional: TIS Procedure:

1970   1980

Look at it.   Save and Process   **Work with it!**
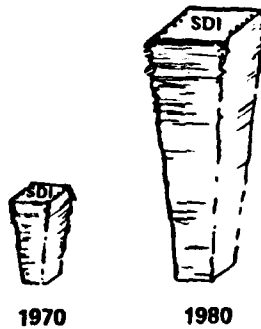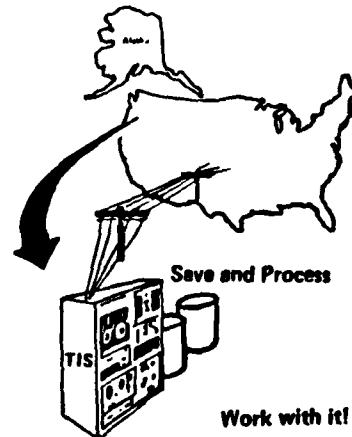
Approximately 100 citations with abstracts can be extracted and saved in 10 minutes at 1200 baud by asynchronous telephone dial-up. Faster transmission is possible with 9600-baud dedicated and conditioned synchronous lines. Second, the DIAL command permits extraction and downloading into one, or more individual, user-named files that can be opened and closed at liberty by special control characters during a session, e.g.

ESCAPE CTRL-A —   prompts the user for a file name and saves the viewed information therein. An additional ESC-CTRL-A closes the file. If the file already exists, the new information is appended to permit progressive creation of a cumulative subject datafile.

ESCAPE CTRL-B —   sends a local file from TIS to a remote machine. This has particular importance when downloaded and saved information is to be transferred to more powerful computers for analysis, or is to be shared with someone else via TIS.

Other special control characters permit the user to stop the viewing, and/or saving, of information and to address the local or remote computer selectively.

The legal and monetary implications of downloading and sharing information extracted from other centers must be considered carefully.

3. POST-PROCESSING OF BIBLIOGRAPHIC INFORMATION

When a retrospective search is carried out for a new field of interdisciplinary research, it is not unusual to obtain thousands of citations from different information vendors, in different formats, with redundant overlap.

Recently, a request made to the Transportation Systems Research program at LLNL to identify foreign R&D in electric batteries yielded 1.5 ft of printouts from federal and commercial information vendors. Carried out by conventional means, the 2000 citations purchased at a cost of about $1700 were quite useless. It was very difficult to convey to DOE headquarters meaningful statistics, insights, or magnitudes of ongoing R&D abroad in time. It took six days alone for the off-line prints to arrive by mail. One solution to this problem, at least for databases from DOE/RECON, is the post-processing of downloaded citations with TIS.

TIS offers programs for the archival diskstorage of retrieved information in a convenient, user-defined filing system. This permits results to be organized and aggregated in a suitable manner. Redundant citations can be eliminated by their congruent main data fields, primarily by comparison of authors and titles. The resulting unique set is then reviewed and analyzed online by self-guided routines:

* REVIEW	citations for relevancy.
* DISPLAY	graphs of publication rates.
* PERMUTE	multi-term expressions in data fields.
* CROSS-CORRELATE	contents of data fields, with statistics.
* CONCORD	citations by author, subject, descriptors, etc.

The REVIEW command permits online determination of relevancy. Citations are shown on the screen reformatted by accentuation and indentation of titles, authors, and abstracts. This renders them more readable than citations commonly offerred by information centers. The viewer may keep or discard any citation shown and assign his own category and relevancy code. He may add comments, order the full-length text, and define and fill new data fields for numeric and/or administrative purposes. Retained and annotated citations are saved in new user-named files. Fields defined during the review process can subsequently be used with other fields for post-processing.

ON-LINE DETERMINATION OF RELEVANT SET:

Relevancy?	0 - 9, X,
IIS-Category?	ABC
Order Text?	Y/N
Comment?	

LiF
$\lambda = 23.5$

reconreview
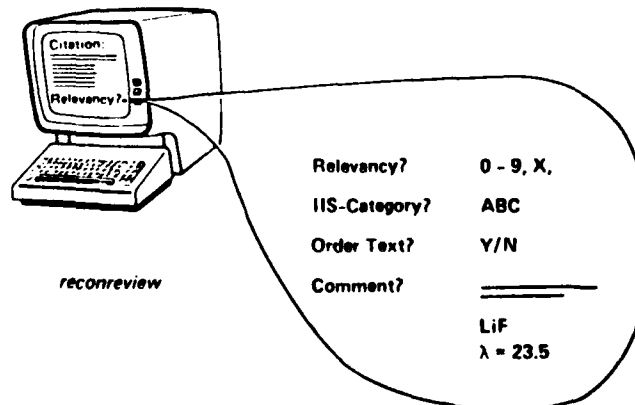
The DISPLAY of the publication rate for a particular search topic, an institute, or author provides an immediate indication of the effort or growth in their field of activity. It is probably the first display an end-user may wish to see and is carried out on TIS by the simple command YEAR-GRAPH. In most cases, as shown in the example below, we see an apparent decline after a sharp rise in the publication rate. This decline is predominantly the lag in time between the publication of the primary literature and its entry into the secondary online database holdings. To appreciate the above average increase in publications in a particular field, one has to compare it with the total annual increase in the publication rate. In the sciences, this rate is now about 13%.[6]

DOE/RECON Data Analysis Program
Number of Publications per Year
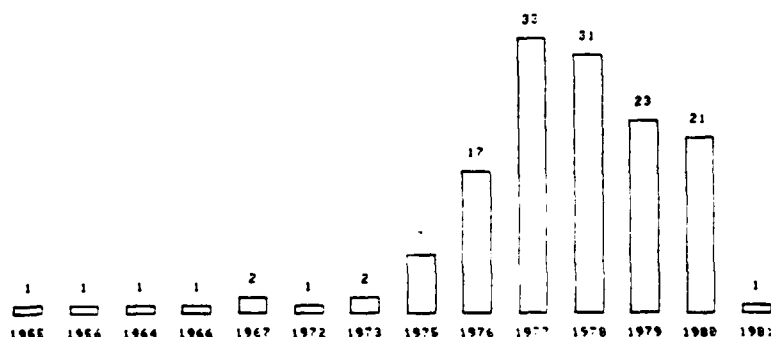
The most significant aspect of post-processing is probably the time-dependent change, or momentum, of a particular field in R&D, derived from the statistics of its permuted descriptive terms. The PERMUTE command of our post-processing routines provides this option by counting the number of times a specified term appears in the message-carrying fields of citations, like the title, abstract, descriptor, category, etc. This is done by analyzing single and compound expressions containing up to four terms [e.g., solar energy conversion experiments]. All compound expressions of this type appearing in the selected data fields are presented to the viewer online, alphabetically ordered, with their frequency of occurence. The tables following show this for two recent projects carried out by the Research Information Group of the Technical Information Department at LLNL for DOE patents and the DOE flywheel program.

Analysis of Descriptors for DOE Patens.

141	DESIGN
3F	FABRICATION
17	RADIOACTIVE WASTE PROCESSING
17	WATER TYPE REACTORS
16	CONTAINERS
16	CHEMICAL PREPARATION
14	HIGH TEMPERATURE
14	CONFIGURATION
13	SPECIFICATIONS
13	ELECTRONIC CIRCUITS
12	CHEMICAL REACTIONS
11	URANIUM
11	GASES
11	EQUIPMENT
10	PREPROCESSING
9	WATER
9	SOLIDIFICATION
9	SAFETY
9	PRODUCTION
9	LIQUID WASTES
9	COAL
9	CHEMICAL EXPLOSIVES
8	EXPORTS
8	SEPARATION PROCESSES
8	REACTOR CORES
8	PRESSURE DEPENDENCE
8	POWDERS
8	LASER RADIATION
	etc.

Analysis of Descriptors & Titles for Flywheel

573	FLYWHEEL
505	ENERGY
303	STORAGE
278	SYSTEM
230	ENERGY STORAGE
178	POWER
174	DESIGN
162	VEHICLE
137	SYSTEM
135	ELECTRIC
125	FLYWHEEL
104	COMPOSITE
102	ROTOR
93	PROGRAM
91	PERFORMANCE
90	MOTOR
	CONTROL
	RESULTS
	etc.

CROSS-CORRELATIONS of expressions between any two fields of the citations can provide new insight. For example, by cross-correlating authors, we can see at a glance who is working with whom. A cross-correlation of the author field with the descriptor field shows, in alphabetic order, the statistics of indexing terms assigned to the work of a particular person for his entire professional career. When carried out in yearly increments, this routine can be used to judge the change of emphasis with time.

Barlow, T.M.

11*	Barlow, T.M.
1	Burrows, C.R.
2	Chiao, T.T.
2	Cornell, E.P.
3	Crothers, W.T.
2	Frank, D.N.
3	Kulkarni, S.V.
1	Reimers, E.
2	Rinde, J.A.
2	Turnbull, F.G.
31	TOTAL

Bauer, W.H.

2*	Bauer, W.H.
1	Brobeck, W.M.
1	Younger, F.C.
4	TOTAL

Beachley, N.H.

10*	Beachley, N.H.
1	Dietrich, A.
9	Frank, A.A.
1	Harter, R.
1	Jamzadeh, F.
1	Lau, K.
1	Otis, D.R.
1	Stockman, D
1	Volz, T.
26	TOTAL

Davis, D.

1	COMMERCIALIZATION
2	FLYWHEEL ENERGY STORAGE
	FLYWHEEL-POWERED VEHICLES
2	FLYWHEELS
	HYBRID ELECTRIC-POWERED VEHICLES
1	OPERATION
2	PERFORMANCE TESTING
3	PERFORMANCE
25	TOTAL

General Electric Co., Schenectady, NY(USA) Corporate Research and Development Dept.

	COMPUTER CALCULATIONS
	COMPUTERIZED SIMULATIONS
2	CONTROL EQUIPMENT
	COST BENEFIT ANALYSIS
2	DESIGN
	ELECTRIC GENERATORS
	ENERGY STORAGE SYSTEMS
2	EXPERIMENTAL DATA
2	FABRICATIONS
	FEASIBILITY STUDIES
	FLYWHEEL ENERGY STORAGE
4	FLYWHEEL POWERED VEHICLES
1	FLYWHEELS
	HYBRID ELECTRIC-POWERED VEHICLES
	LIFE CYCLE COST
2	PERFORMANCE TESTING
	RESEARCH PROGRAMS
	STEELS
	WELDING
34	TOTAL

Author-author correlation.   Author-descriptor correlation

CONCORDANCES generated by author, descriptor, corporate author, or country yield succinct listings of bibliographic citations in a particular field. These alphabetical indexes are similar to those commonly produced as look-up tables for authors or subjects. In this case, they are created at the pleasure of the user, online, on the contents of any citation field.

Erdal, B.R.

| 1975 | Grant, P.M., Erdal, B.R., O'Brien, H.A. | RUBIDIUM RADIOISOTOPE GENERATOR | 77P0069342 |
| 1976 | Grant, P.M.; Erdal, B.R., O'Brien, H.A. | Sr-Rb RADIOISOTOPE GENERATOR | 77P0019544 |

Essebaggers, J.

| 1975 | Essebaggers, J. | ENERGY ABSORBER FOR SODIUM-HEATED HEAT EXCHANGER | 76P0058174 |

Evans, H.W.

| 1976 | Evans, H.W. | PROTECTIVE AIR LOCK | 76P0076192 |

Everleigh, J.W.

| 1976 | Sartory, W.K.; Everleigh, J.W. | CENTRIFUGE APPARATUS | 77P0007259 |

Farnum, E.H.

| 1975 | Farnum, E.H., Fries, R.J. | METHOD FOR SIZING HOLLOW MICROSPHERES | 77P0040417 |
| 1976 | Farnum, E.H.; Fries, R.J. | METHOD FOR NONDESTRUCTIVE FUEL ASSAY OF LASER FUSION TARGETS | 77P0085122 |

Farrior, W.L. Jr.

| 1976 | Farrior, W.L. Jr. | REGENERABLE SORBENT AND METHOD FOR REMOVING HYDROGEN SULFIDE FROM HOT GASEOUS MIXTURES | 77P0085122 |

COAL DEPOSITS

1976	Archibald, P.B.	EXPLOSIVE FLUID TRANSMITTED SHOCK METHOD FOR MINING DEEPLY BURIED COAL	77P0005251
1976	Fisher, S.T.; Fisher, C.B.	EXTRACTION OF HYDROCARBONS IN SITU FROM UNDERGROUND HYDROCARBON DEPOSITS	77P0005129
1976	Pasini, J. III; Overbey, W.K. Jr.	METHOD FOR REMOVAL OF METHANE FROM COALBEDS	76P0063102

COAL FINES

| 1976 | Coates, R.L. | GASIFICATION OF CARBONACEOUS SOLIDS | 77P0085185 |

COAL GASIFICATION

1976	Coates, R.L.	GASIFICATION OF CARBONACEOUS SOLIDS	77P0085185
1976	Donath, E.E.	METHOD AND APPARATUS FOR REMOVING COARSE UNENTRAINED CHAR PARTICLES FROM THE SECOND STAGE OF A TWO-STAGE COAL GASIFIER	76P0062836
1976	Fisher, S.T.; Fisher, C.B.	EXTRACTION OF HYDROCARBONS IN SITU FROM UNDERGROUND HYDROCARBON DEPOSITS	77P0005129
1976	Komar, C.A.	METHOD FOR CONTROL OF SUBSURFACE COAL GASIFICATION	77P0085189
1976	Lee, R.S	OXIDATION OF COAL WATER SLURRY FEED TO HYDROGASIFIER	77P0005121

Extracts from concordances of DOE patents by authors and descriptors.

Output from the post-processing routines can be saved in files for subsequent use, or for transfer to electronic word processors and merging with reports, or can be sent to typesetters as camera-ready copy for publication. A full description of present TIS post-processing capabilities has been published.[7] Similar approaches are being developed elsewhere.[8]

5. CONCLUSION

Downloading and post-processing of bibliographic citations and numeric data offer the information specialist powerful and cost-effective tools for the repackaging of search results and the delivery of more relevant information products. At present, our work at LLNL is concentrated on applying and refining these tools for DOE/RECON and is being sponsored by the Department of Energy Technical Information Center (DOE/TIC). However, the response has been so favorable that we have been asked to explore the possibilities of extending these capabilities to databases of other federal information systems. This extension requires, where possible, unified command languages and the reformatting of retrieved citations.

These TIS capabilities have been demonstrated by the NASA Industrial Applications Center at the University of Southern California, where information specialists linked their terminals via TIS to clients' terminals elsewhere in the country while conducting search and post-processing with databases from DOE/RECON, and NASA/RECON.[9,10] This lingkage provided simultaneous viewing and voice communication, and instant delivery of the refined product to the end-user, thereby speeding substantially the timely delivery of information products.

Downloading and postprocessing of bibliographic information is being developed at LLNL to improve the transfer of government-sponsored technology among federal agencies and to industry.[11,12] Similar post-processing routines are being developed throughout the information industry. Commercial database producers and information vendors must arrive at practical solutions for the use of such technological innovations.[13]

6. REFERENCES

[1] Cuadra, R.N., Abels, D.M., and Wagner, J., Directory of Online Databases, Quarterly Editions, Vol.3, No.4, 1982, Cuadra Associates, Inc., 2001 Wilshire Blvd., Suite 305, Santa Monica, CA 90403.

[2] Hampel V.E., Gallo, L.E., Kawin R.A., Kopytoff V., McGrogan S.K., O'Connell L.G., Pavel G., Schriebman J.A., and Swanson J.E., First DOE/LLL Workshop on the Technical Management Information System (TMIS), CONF-791258, 1979.

[3] Hampel V.E., Scott W.S., Gallo L.E., Kawin R.A., Kopytoff V., Mallon B., McGrogan S.K., Pavel G., Rabe W.G., Schriebman J.A.,and Swanson J.E., User's Manual for the Technology Information System, Second Edition, M-0112, 1981.

[4] Hampel, V.E., "Fact Retrieval for the 1980s", Proceedings of the Technical Information Panel (TIP), Conference of the Advisory Group for Aerospace Research and Development (AGARD), (North Atlantic Treaty Organization (NATO), Munich, West Germany, 1981), published as AGARD-CPP-304, and UCRL-85749.

[5] Huleatt R.S., Data Downloading: A Report on Who Does What, to Whom, Online '81 Conference Proceedings, 1981, pp. 173.

[6] Hall J.L., and Brown M.J., Online Bibliographic Databases, An International Directory, 2nd Edition, 1981.

[7] Hampel V.E., Schriebman J.A., and Kawin R.A., Post-Processing of Bibliographic Information, September 1981.

[8] Riley C., Bell M., Use of an Intelligent Terminal for Sophisticated Searching Techniques and for the Production of Report Style Searches Online '81 Conference Proceedings, pp. 376, November 1981.

[9] Schriebman, J.A., and Swanson, J.E., Crosslinking of Computer Terminals for Nationwide Tutorial Instruction, UCRL-52964, November 1980.

[10] Wallace, R.S., Hampel, V.E., Mixer, R., Linking the Information Specialist to the End-User for Joint Audio-Visual Searching of the Literature and Immediate Delivery of Search Results, In print, September, 1982.

[11] Hampel, V.E., Kawin, R.A., Lann, N.A., and Scott, W.S., TIS—An Interactive System for Technology Transfer, 7th International Symposium of the Technology Transfer Society, Pasdadena, CA, June 13-14, 1982, UCRL-87703.

[12] Hampel, V.E., Intelligent Gateway Computers for Nationwide Information Systems of the Future, Post Conference Proceedings of the World Future Conference, Washington, D.C., July 19-21, 1982.

[13] Burnam, J.K., Users' Needs for Repacking and Reuse of Information, Information Services and Use, Vol.1, pp. 359-366, North Holland Pub. Co., 1982.

END

DATE
FILMED

9 — 83

DTIC